|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  21st Meeting: Warsaw, PL, 19–26 June 2015 | Document: JCTVC-U0041 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **On 4:4:4 to 4:2:0 conversion - Performance of downsampling and upsampling filters, and MinMax and Closed Loop Filtering** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Information | | |
| *Author(s) or Contact(s):* | A.M. Tourapis, Y. Su, David Singer  1 Infinite Loop Cupertino, CA 95014 USA | Tel: Email: | +1-408-228-7983 atourapis@apple.com |
|  | J. Sole, D. Rusanovskyy, S. Lee, D. Bugdayci, A. Ramasubramonian, M. Karczewicz  5775 Morehouse Drive San Diego, CA 92121-1714 USA |  |  |
|  | C. Fogg |  |  |
|  | A. Duenas, F. Bossen |  |  |
| *Source:* | Apple Inc., Qualcomm Inc., Movielabs Inc., NGCodec Inc. | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This contribution tries to address some issues on chroma format conversion as they were observed within the MPEG HDR/WCG activity. It is suggested that these issues do not only impact HDR/WCG material but also the conversion of SDR material. Considerable benefits, both objective and subjective, could be achieved using improved conversion techniques, as well as techniques that account for the characteristics and the representation of the material.

# Introduction

MPEG recently released a Call for Evidence (CfE) document requesting the submission of HDR/WCG technologies that could demonstrate improved performance over the performance of HEVC Main 10 [1]. evidence would be compared versus anchor bitstreams generated using the official HM software and material converted in what was believed to be the primary format for content distribution using existing technologies. In particular, the material were converted to the Non-Constant Luminance Y’CbCr 4:2:0 format, using the BT.2020 color primaries [2], and the PQ ST 2084 [3]transfer function.

The anchor material were converted to the Non-Constant Luminance Y’CbCr 4:2:0 format by utilizing rather conservative down-sampling and up-sampling methodologies. This was done to avoid or minimize the possibility of conversion artifacts that were observed earlier during this activity. More specifically, the downsampling filters were specified in Section B.1.5.5 and the upsampling filters in Section B.1.5.6 of [4]. These filters are presented below. It should be noted that unlike the recommendation in the BT.2020 specification [2] of using chroma sampling location 2, chroma sampling location 0 is used instead.

Table 1‑1. Downsampling filters used for the MPEG HDR/WCG CfE Anchors

|  |  |  |
| --- | --- | --- |
| Phase k | Coefs c1[k]  4:4:4 🡪 4:2:2  (phase=0) | Coefs c2[k]  4:2:2 🡪 4:2:0  (phase=0.5) |
| -1 | 1 | 0 |
| 0 | 6 | 4 |
| 1 | 1 | 4 |

Table 1‑2. Upsampling filters used for the MPEG HDR/WCG CfE Anchors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Phase | -2 | -1 | 0 | 1 |
| Coef c[k] | -4 | 36 | 36 | -4 |
| Coef d0[k] | -2 | 16 | 54 | -4 |
| Coef d1[k] | -4 | 54 | 16 | -2 |

The magnitude and phase responses of these filters are also shown below:



Figure 1‑1. Magnitude and Phase responses of the [1 6 1]/8 filter



Figure 1‑2. Magnitude and Phase responses of the [1 1]/2 filter



Figure 1‑3. Magnitude and Phase responses of the [-1 9 9 -1]/16 filter

The conversion performance of using these filters, and without compression, for the content used for the CfE can be seen in Table 1‑3.

Table 1‑3. Conversion performance of the MPEG HDR/WCG CfE Anchors

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **tPSNR X** | **tPSNR Y** | **tPSNR Z** | **tPSNR XYZ** | **deltaE psnr** | **mPsnr** |
| *FireEaterClip4000r1* | 58.59 | 64.34 | 53.13 | 56.56 | 48.28 | 43.72 |
| *Tibul2Clip4000r1* | 55.24 | 60.09 | 52.72 | 55.07 | 45.46 | 49.07 |
| *Market3Clip4000r2* | 53.46 | 62.01 | 46.22 | 50.14 | 36.66 | 46.95 |
| *AutoWelding* | 48.20 | 58.57 | 40.99 | 44.93 | 39.55 | 31.62 |
| *BikeSparklers* | 45.96 | 58.45 | 35.82 | 40.15 | 33.97 | 34.47 |
| *ShowGirl2Teaser* | 52.14 | 62.98 | 44.89 | 48.84 | 39.11 | 37.63 |
| *StEM\_MagicHour* | 50.69 | 62.67 | 38.74 | 43.22 | 36.02 | 36.22 |
| *StEM\_WarmNight* | 51.60 | 61.52 | 39.13 | 43.63 | 36.71 | 35.28 |
| *BalloonFestival* | 52.60 | 61.89 | 45.33 | 49.27 | 40.33 | 46.74 |

Other downsampling filters [5] were discussed in earlier stages of the activity, however it was claimed that these could result in artefacts, especially near edges of considerable brightness difference. These artefacts seemed to stem from the use of negative coefficients (Figure 1‑4) in these filters. A closer examination of these filters, especially when applied on integer data, is that these filters, given the negative coefficients, can potentially result in values much larger or much smaller than the original values within a particular neighbourhood. This makes it easier for such filters to exceed the valid chroma range, as also discussed in [8]. Although this problem can also happen in SDR content, it seems to be much more accented in HDR material given the more significant differences in dynamic range, as well as the behaviour of the PQ transfer function[[1]](#footnote-1).



Figure 1‑4. Coefficients of the downsampling filter (phase 0) in [5]



Figure 1‑5. Example behavior of different filters

The magnitude and phase responses of the filters in [5] are shown in Figure 1‑6.



Figure 1‑6. Magnitude and Phase responses for the phase 0 downsampling filter in [5]

In this document we identify solutions for improving downconversion and upconversion, therefore resulting in considerably better conversion behaviour, both in terms of objective and subjective performance, than that observed for the anchor content.

# Improved Up-sampling

The current anchor generation process, as mentioned in the previous section, uses a rather simple filter, i.e. a Lanczos-2, for the up-sampling process. The Lanczos class of filters involves essentially windowed sinc functions and is very commonly used for up-conversion. The Lanczos-2 is the simplest form of these filters and it is found in several lower-end applications/system, and is preferred against other simple filters such as the bilinear filter. It is well known that moving to a higher order Lanczos filter would commonly improve interpolation performance since this allows the Lanczos filter to better approximate the Sinc filter.

Given that, we explored how performance would improve by using higher order lanczos filters during up‑sampling. We have implemented Lanczos filters with order 3 up to 6 in our extended implementation of HDRTools, which we could make available to anyone interested after the meeting. The Magnitude and Phase responses of the Lanczos-4 filter are shown in Figure 2‑1. The coefficients of this filter are shown in Figure 2‑2.



Figure 2‑1. Magnitude and Phase responses for the phase 0 Lanczos-4 filter



Figure 2‑2. Coefficients of the Lanczos-4 filter (phase 0)

Although we will not be presenting these results here, we have found that using the Lanczos-4 filter provides a better compromise between improved sharpness and reduced aliasing/ringing after the interpolation is completed. This is especially true when converting compressed content, since those may have been impacted by quantization noise and especially blockiness.

# Floating point/High Precision filtering

Figure 3‑1 shows the end to end coding and decoding chain used for the anchor generation process. In this process it is assumed that conversion from 444 to 420 sampling, and vice versa, should be performed using the 10-bit quantized data.

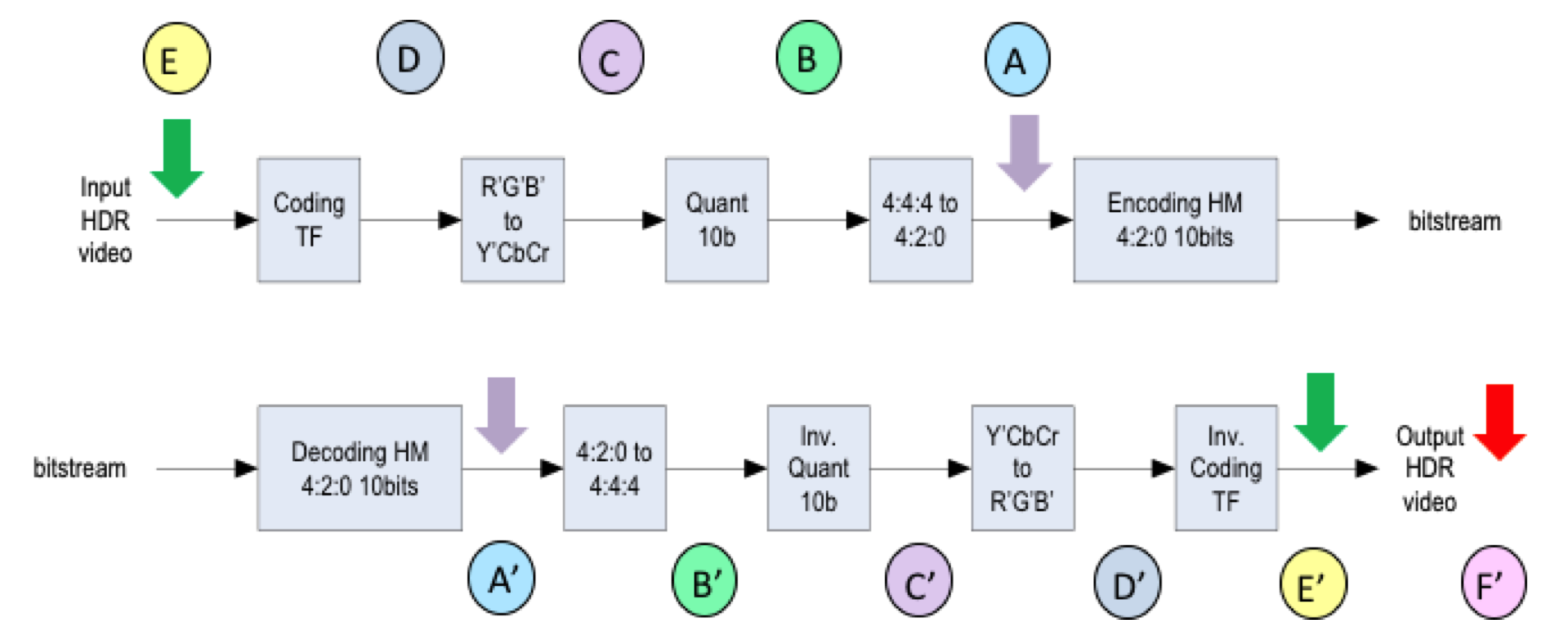


Figure 3‑1. End to end “Anchor” coding and decoding chain

However, there is nothing that says that down and upconversion needs to follow this exact workflow. In particular, one may claim that performing the chroma downconversion using the 10-bit quantized data is quite suboptimal since now the filter has to operate on data that contain quantization errors. Instead one may prefer to perform down or up-conversion in a much higher precision domain, e.g. directly on floating precision data, thus avoiding or limiting propagating such quantization errors. This process is shown in Figure 3‑2.



Figure 3‑2. Alternative end to end coding and decoding chain with floating point conversion

Another alternative down-conversion process is shown in Figure 3‑3. In this scenario, downconversion is performed in the linear domain and directly on RGB data. However, and especially for the NCL Y’CbCr process, this method is much more complicated to perform, mainly in terms of upconversion.



Figure 3‑3. Encoding and decoding chain using linear conversion

# MinMax Filtering

To reduce if not eliminate the problem with the down-conversion process that we identified in Section 1, we introduce the concept of *MinMax* filtering when performing filtering for down-conversion from 4:4:4 to 4:2:0. In particular the process involves clipping the output of the filtering process given a minimum and maximum limit that are derived dynamically for every sample position. In particular, we commonly set these limits as the smallest and largest value respectively within a small neighborhood around the current sample that is being filtered. In a particular example, we consider only the current sample and its two neighbors according to the direction of the filtering process, for this limiting process.

# Closed Loop Color Conversion

In many applications color conversion from R’G’B’ data to Y’CbCr is performed using a 3x3 matrix conversion. That is

However, Cb and Cr are also related to Y’ in the following manner:

with

with

Given these formulations, instead of performing the conversion as a single 3x3 matrix conversion process, one could use a two step approach, where one first generates the Y’ and, given Y’, then generate Cb and Cr. The trick, however, in this technique is also account for the Y’ quantization in this conversion. More specifically, if we know that Y’ will be quantized to a particular precision Q, the conversion process can be performed as:

This conversion process, which we will refer to as the “Closed Loop Color Conversion” method, allows us to account for luma quantization during the color conversion process. Assuming no subsequent conversion, this enables us to get a more accurate reconstruction for the B’ and R’ data after decoding.

# Performance Results

In [9] two different responses to the MPEG HDR/WCG CfE were presented. In one of the methods, a very similar method to the ones used for the anchors was used, by basically utilizing the NCL representation while applying a scaling function onto the chroma data prior to quantization. Furthermore, the data were quantized to the SDI video range instead of the Limited range, allowing higher precision for the representation. No further modifications to the down and up-conversions were performed. The conversion performance of that scheme, without any encoding considered, is shown in Table 6‑1. In Table 6‑2 we present the conversion results of the same scheme but with the addition of the Lanczos 4, instead of Lanczos 2, upconversion. Table 6‑3 also includes the floating precision conversion as well as the chroma clipping results, whereas in Table 6‑4 we also include the improved down-conversion techniques, excluding the closed loop conversion, in the process.

Table 6‑1. Conversion performance of the NCL scheme presented in [9]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **tPSNR X** | **tPSNR Y** | **tPSNR Z** | **tPSNR XYZ** | **deltaE psnr** | **mPsnr** |
| *FireEaterClip4000r1* | 59.35 | 65.02 | 53.56 | 57.07 | 48.78 | 44.88 |
| *Tibul2Clip4000r1* | 55.71 | 60.59 | 53.04 | 55.45 | 45.99 | 49.44 |
| *Market3Clip4000r2* | 53.61 | 62.23 | 46.29 | 50.23 | 36.95 | 47.05 |
| *AutoWelding* | 48.28 | 58.70 | 41.03 | 44.98 | 39.64 | 31.68 |
| *BikeSparklers* | 46.01 | 58.54 | 35.84 | 40.17 | 34.01 | 34.49 |
| *ShowGirl2Teaser* | 52.26 | 63.28 | 44.95 | 48.91 | 39.28 | 37.74 |
| *StEM\_MagicHour* | 50.77 | 62.89 | 38.75 | 43.24 | 36.05 | 36.24 |
| *StEM\_WarmNight* | 51.71 | 61.69 | 39.14 | 43.65 | 36.74 | 35.31 |
| *BalloonFestival* | 52.69 | 62.13 | 45.37 | 49.33 | 40.82 | 46.82 |

Table 6‑2. NCL scheme in [9] with Lanczos 4 upconversion

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **tPSNR X** | **tPSNR Y** | **tPSNR Z** | **tPSNR XYZ** | **deltaE psnr** | **mPsnr** |
| *FireEaterClip4000r1* | 59.85 | 65.41 | 54.00 | 57.51 | 48.98 | 45.71 |
| *Tibul2Clip4000r1* | 55.79 | 60.66 | 53.13 | 55.54 | 46.02 | 49.52 |
| *Market3Clip4000r2* | 53.81 | 62.44 | 46.56 | 50.49 | 36.99 | 47.30 |
| *AutoWelding* | 48.48 | 58.97 | 41.21 | 45.16 | 39.74 | 31.77 |
| *BikeSparklers* | 46.11 | 58.78 | 35.86 | 40.21 | 34.03 | 34.51 |
| *ShowGirl2Teaser* | 52.28 | 63.35 | 44.94 | 48.91 | 39.31 | 37.67 |
| *StEM\_MagicHour* | 50.77 | 63.03 | 38.75 | 43.24 | 36.05 | 36.21 |
| *StEM\_WarmNight* | 51.77 | 61.80 | 39.10 | 43.61 | 36.73 | 35.13 |
| *BalloonFestival* | 53.34 | 62.76 | 45.74 | 49.74 | 40.91 | 47.10 |

Table 6‑3. NCL scheme in [9] with floating point Lanczos-4 up conversion and Chroma Clipping [8]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **tPSNR X** | **tPSNR Y** | **tPSNR Z** | **tPSNR XYZ** | **deltaE psnr** | **mPsnr** |
| *FireEaterClip4000r1* | 59.95 | 65.46 | 54.16 | 57.66 | 49.07 | 46.05 |
| *Tibul2Clip4000r1* | 55.85 | 60.71 | 53.24 | 55.62 | 46.14 | 49.59 |
| *Market3Clip4000r2* | 53.83 | 62.44 | 46.59 | 50.51 | 37.05 | 47.32 |
| *AutoWelding* | 48.48 | 58.97 | 41.21 | 45.16 | 39.77 | 31.76 |
| *BikeSparklers* | 46.12 | 58.78 | 35.87 | 40.21 | 34.04 | 34.52 |
| *ShowGirl2Teaser* | 52.36 | 63.58 | 44.96 | 48.94 | 39.36 | 37.38 |
| *StEM\_MagicHour* | 50.79 | 63.04 | 38.76 | 43.25 | 36.06 | 36.22 |
| *StEM\_WarmNight* | 51.77 | 61.80 | 39.10 | 43.62 | 36.74 | 35.14 |
| *BalloonFestival* | 53.35 | 62.79 | 45.75 | 49.75 | 40.99 | 47.14 |

Table 6‑4. NCL scheme in [9] with improved down and up-conversion

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **tPSNR X** | **tPSNR Y** | **tPSNR Z** | **tPSNR XYZ** | **deltaE psnr** | **mPsnr** |
| *FireEaterClip4000r1* | 60.46 | 65.86 | 54.69 | 58.17 | 49.25 | 46.63 |
| *Tibul2Clip4000r1* | 56.27 | 61.10 | 53.68 | 56.05 | 46.31 | 49.93 |
| *Market3Clip4000r2* | 54.32 | 62.87 | 47.10 | 51.02 | 37.08 | 47.74 |
| *AutoWelding* | 49.00 | 59.44 | 41.73 | 45.67 | 39.91 | 32.18 |
| *BikeSparklers* | 46.62 | 59.20 | 36.38 | 40.72 | 34.19 | 34.98 |
| *ShowGirl2Teaser* | 52.81 | 63.87 | 45.40 | 49.39 | 39.52 | 37.75 |
| *StEM\_MagicHour* | 51.27 | 63.44 | 39.26 | 43.75 | 36.25 | 36.64 |
| *StEM\_WarmNight* | 52.21 | 62.17 | 39.54 | 44.06 | 36.92 | 35.51 |
| *BalloonFestival* | 53.94 | 63.25 | 46.28 | 50.29 | 40.97 | 47.45 |

It can be observed that each step can provide considerable improvement over the performance of the scheme in [9]. A plot that demonstrates the tPSNR-Y gains over the anchors is shown in Figure 6‑1. It can be seen from this plot that although for some material the techniques applied in [9] can present some moderate gains, especially, for example in the FireEater and Tibul sequences, much more moderate and consistent gains could be achieved especially if using a better down-sampler (+DOWN). Using also a better upsampler (+LZ4) can in same cases provide additional gains, and as much as 0.63dB (BalloonFestival). Floating point processing and chroma clipping, as expected, provide more “subtle”, but still noticeable, improvements.

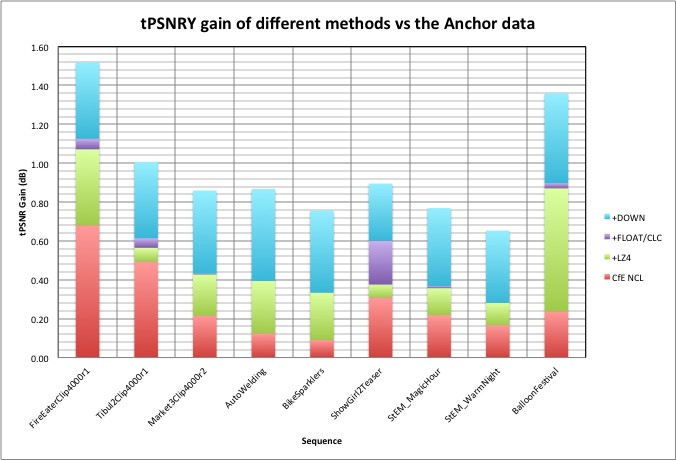


Figure 6‑1. tPSNR-Y gains over the Anchors for different methods

Similar improvements can be observed also on the BT.709 cases. In Table 6‑5 the conversion results by using the anchor approach, and the conversion results by using the improved downsampling and upsampling techniques presented in this document are shown. It should be noted that the additional techniques used in [9] were not used here.

Table 6‑5. Conversion performance of anchors and improved downconversion for BT.709 data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Anchors** | | | | | **Improved Conversion** | | | | |
|  | **tPSNR** | | |  |  | **tPSNR** | | |  |  |
|  | **X** | **Y** | **Z** | **deltaE** | **mPsnr** | **X** | **Y** | **Z** | **deltaE** | **mPsnr** |
| *FireEater* | 54.462 | 57.104 | 54.734 | 47.622 | 42.478 | 55.608 | 58.147 | 55.914 | 48.116 | 44.341 |
| *Tibul2* | 50.229 | 52.179 | 59.633 | 44.373 | 46.489 | 50.826 | 52.774 | 60.365 | 44.713 | 47.012 |
| *Market3* | 47.212 | 49.314 | 44.58 | 36.692 | 42.883 | 47.895 | 50.054 | 45.249 | 36.842 | 43.423 |

Additional results that also account for compression will also be presented during the meeting.

# Conclusion

This contribution presents several techniques that could be utilized for improved down and upconversion of video sequences, especially in the context of 4:4:4 to 4:2:0 color format conversion. Considerably objective, as well as subjective, improvements are observed when utilizing these methods versus the techniques used in [4].

# References

1. ISO/IEC 230082:2014 HEVC Second Edition
2. Recommendation ITU-R BT. 2020, “Parameter values for ultra-high definition television systems for production and international programme exchange” (2012).
3. SMPTE ST 2084:2014, “High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays”, 2014.
4. Call for Evidence (CfE) for HDR and WCG Video Coding, MPEG document M36131, Switzerland, Feb. 2015
5. G. Sullivan, “Color format downconversion for test sequence generation,” JVT document JVT-I018, San Diego, USA, Sep. 2003
6. G. Sullivan, “Color format upconversion for video display,” JVT document JVT-I019, San Diego, USA, Sep. 2003
7. S. Sun and J. Reichel, “AHG Report on Spatial Scalability Resampling,” JVT document JVT-R006, Bangkok, Thailand, Jan. 2006
8. A.M. Tourapis, Y. Su, D. Singer, J. Sole, D. Rusanovskyy, M. Karczewicz, C. Fogg, A. Duenas, and F. Bossen, “Clipping during 4:2:0 to 4:4:4 conversions”, JCT-VC/MPEG document JCTVC-U0042/m36272, Warsaw, Poland, June 2015
9. D. Rusanovskyy, S. Lee, D. Bugdayci, A. Ramasubramonian, J. Sole, M. Karczewicz, A.M. Tourapis, Y. Su, D. Singer, C. Fogg, A. Duenas, and F. Bossen, “Single layer non-normative (category 3a) NCL and CL responses to the Call for Evidence on HDR/WCG”, MPEG Document M36256, Warsaw, Poland, June 2015

# Patent rights declaration(s)

**Apple Inc may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**

**Qualcomm Inc may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**

**Movielabs Inc may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**

**NGCodec Inc may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**

1. Chroma downsampling and upsampling is commonly applied in the TF domain. [↑](#footnote-ref-1)